

# **Sequence Homology Search**

SEQ ID No 9

AC015701  
 LOCUS AC015701 174149 bp DNA HTG 07-APR-2000  
 DEFINITION Homo sapiens chromosome 11 clone RP11-210K21 map 11, WORKING DRAFT  
 SEQUENCE, 24 unordered pieces.  
 ACCESSION AC015701  
 VERSION AC015701.3 GI:7523736  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 174149)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 TITLE Homo sapiens chromosome 11, clone RP11-210K21  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 174149)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
 Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,  
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,  
 Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
 Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,  
 Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,  
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,  
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT On Apr 7, 2000 this sequence version replaced gi:6479169.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: L1373  
 Center clone name: 210\_K\_21  
 ----- Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 144096 bases at least Q40  
 Consensus quality: 161320 bases at least Q30  
 Consensus quality: 168048 bases at least Q20  
 Insert size: 187000; agarose-fp  
 Insert size: 171849; sum-of-contigs  
 Quality coverage: 3.5 in Q20 bases; agarose-fp  
 Quality coverage: 3.8 in Q20 bases; sum-of-contigs  
 -----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 24 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```
*      1      1827: contig of 1827 bp in length
*      1828 1927: gap of      100 bp
*      1928      3978: contig of 2051 bp in length
*      3979 4078: gap of      100 bp
*      4079      6795: contig of 2717 bp in length
*      6796 6895: gap of      100 bp
*      6896      9683: contig of 2788 bp in length
*      9684 9783: gap of      100 bp
*      9784      11884: contig of 2101 bp in length
*     11885 11984: gap of      100 bp
*     11985      16473: contig of 4489 bp in length
*     16474 16573: gap of      100 bp
*     16574      20142: contig of 3569 bp in length
*     20143 20242: gap of      100 bp
*     20243      24886: contig of 4644 bp in length
*     24887 24986: gap of      100 bp
*     24987      29158: contig of 4172 bp in length
*     29159 29258: gap of      100 bp
*     29259      33784: contig of 4526 bp in length
*     33785 33884: gap of      100 bp
*     33885      38874: contig of 4990 bp in length
*     38875 38974: gap of      100 bp
*     38975      44050: contig of 5076 bp in length
*     44051 44150: gap of      100 bp
*     44151      49356: contig of 5206 bp in length
*     49357 49456: gap of      100 bp
*     49457      54065: contig of 4609 bp in length
*     54066 54165: gap of      100 bp
*     54166      60407: contig of 6242 bp in length
*     60408 60507: gap of      100 bp
*     60508      68554: contig of 8047 bp in length
*     68555 68654: gap of      100 bp
*     68655      77303: contig of 8649 bp in length
*     77304 77403: gap of      100 bp
*     77404      88452: contig of 11049 bp in length
*     88453 88552: gap of      100 bp
*     88553      97198: contig of 8646 bp in length
*     97199 97298: gap of      100 bp
*     97299      111202: contig of 13904 bp in length
*    111203 111302: gap of      100 bp
*    111303      122947: contig of 11645 bp in length
*    122948 123047: gap of      100 bp
*    123048      137344: contig of 14297 bp in length
*    137345 137444: gap of      100 bp
*    137445      150883: contig of 13439 bp in length
*    150884 150983: gap of      100 bp
*    150984      174149: contig of 23166 bp in length.
```

```
FEATURES
  source      Location/Qualifiers
              1. .174149
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="11"
              /map="11"
```

```

/clone="RP11-210K21"
/clone_lib="RPCI-11 Human Male BAC"
misc_feature 1. .1827
/note="assembly_fragment"
misc_feature 1928. .3978
/note="assembly_fragment"
misc_feature 4079. .6795
/note="assembly_fragment"
misc_feature 6896. .9683
/note="assembly_fragment"
misc_feature 9784. .11884
/note="assembly_fragment"
misc_feature 11985. .16473
/note="assembly_fragment"
misc_feature 16574. .20142
/note="assembly_fragment"
misc_feature 20243. .24886
/note="assembly_fragment"
misc_feature 24987. .29158
/note="assembly_fragment"
misc_feature 29259. .33784
/note="assembly_fragment"
misc_feature 33885. .38874
/note="assembly_fragment"
misc_feature 38975. .44050
/note="assembly_fragment"
misc_feature 44151. .49356
/note="assembly_fragment"
misc_feature 49457. .54065
/note="assembly_fragment"
misc_feature 54166. .60407
/note="assembly_fragment"
misc_feature 60508. .68554
/note="assembly_fragment
clone_end:T7
vector_side:right"
misc_feature 68655. .77303
/note="assembly_fragment"
misc_feature 77404. .88452
/note="assembly_fragment"
misc_feature 88553. .97198
/note="assembly_fragment"
misc_feature 97299. .111202
/note="assembly_fragment"
misc_feature 111303. .122947
/note="assembly_fragment
clone_end:SP6
vector_side:left"
misc_feature 123048. .137344
/note="assembly_fragment"
misc_feature 137445. .150883
/note="assembly_fragment"
misc_feature 150984. .174149
/note="assembly_fragment"

```

BASE COUNT 53260 a 32685 c 32782 g 53122 t 2300 others  
 ORIGIN

Query Match 49.0%; Score 83.8; DB 73; Length 174149;  
 Best Local Similarity 70.8%; Pred. No. 4.5e-15;

Matches 114; Conservative 0; Mismatches 46; Indels 1; Gaps 1;

```
Qy      4  cacanangannggncntgtgaggacacagcnagaagcaagtctntgcatgncnagaagaa 63
      |||| | || || | || ||||| |||| | | | | | | |||
Db 5282 CACAGAAGAGAGGCCATGGGAGGACACAGAGAGAAGGTGGTGTCTACAAGCCGAGGAGAG 5341

Qy      64  cggcctcaacagacaccanncctgccagcaccttgatcttggcttntggcctccagaact 123
      ||| || |||| |||| ||||| ||||| ||||| | ||||| |||||
Db 5342 AGGCTGCACCAGAAACCAACCCTGCCAGCAGTTTGATCTTGGACTTCAGCCTCCAGAACT 5401

Qy     124  gtgaaagantaaagattctgttgtttaagccagtacaaaat 164
      |||| | | |||| ||||| ||||| | | ||| ||
Db 5402 GTGAGAAAATAAA-CTTCTGTTGTTTAGGTCCCCCAGCAT 5441
```

Q29033/c

Query Match 32.4%; Score 55.4; DB 1; Length 279;  
Best Local Similarity 67.3%; Pred. No. 5e-10;  
Matches 99; Conservative 0; Mismatches 43; Indels 5; Gaps 2;

SEQ ID NO 15

```

LOCUS      AQ534984      478 bp      DNA      GSS      18-MAY-1999
DEFINITION RPCI-11-380E18.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-
            380E18, genomic survey sequence.
ACCESSION  AQ534984

VERSION    AQ534984.1   GI:4846674
KEYWORDS   GSS.
SOURCE     human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1   (bases 1 to 478)

```

AUTHORS Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C.

TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building

JOURNAL Unpublished (1997)

COMMENT On Dec 15, 1999 this sequence version replaced gi:4214464.  
Other\_GSSs: RPCI-11-380E18.TV  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@tigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs ([info@resgen.com](mailto:info@resgen.com)). BAC end search page: [http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html).  
Seq primer: SP6  
Class: BAC ends.

FEATURES Location/Qualifiers

source 1. .478  
/organism="Homo sapiens"  
/db\_xref="GDB:7645649"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-380E18"  
/clone\_lib="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI; RPCI11 Human Male BAC Library"

BASE COUNT 149 a 68 c 130 g 131 t

ORIGIN

Query Match 99.0%; Score 164.4; DB 105; Length 478;  
Best Local Similarity 99.4%; Pred. No. 4.6e-42;  
Matches 165; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy      1 tcagtatcctgacctggcaaggtgttccttaacctcccctctggatcccccttagcacac 60
        |||
Db      460 TCAGTATCCTGACCTGGCAAGGTGTTCTTAACCTCCCCTCTGGATCCCCCTTAGCACAC 401

Qy      61 atctgggacaatggagcgttcagcaccacggacagcattacaccctcttcaagtgtgt 120
        |||
Db      400 ATCTGGGACAATGGAGCGTTCAGCACACGGACAGCATTACACCCTCTTCAAGTGCTTGT 341

Qy      121 taaggccatttgtctattttcactctcaagtaaataaaaaatattttt 166
        |||
Db      340 TAAAGCCATTTGTCTATTTCACTCTCAAGTAAATAAAAATATTTT 295

```

SEQ ID NO: 18

T03943/c

ID T03943 standard; DNA; 4823 BP.

AC T03943;

DT 29-APR-1996 (first entry)  
 DE Human thrombopoietin genomic coding sequence.  
 KW Thrombopoietin; erythropoiesis stimulator;  
 KW haematopoietic polypeptide; treatment; thrombocytopenia; anaemia;  
 KW ds.  
 OS Homo sapiens.  
 PN WO9521626-A1.  
 PD 17-AUG-1995.  
 PF 09-FEB-1995; U01829.  
 PR 14-FEB-1994; US-196025.  
 PR 25-FEB-1994; US-203197.  
 PR 21-MAR-1994; US-215203.  
 PR 01-JUN-1994; US-252491.  
 PR 09-AUG-1994; US-288417.  
 PR 07-NOV-1994; US-335566.  
 PR 01-DEC-1994; US-347748.  
 PA (UNIW ) UNIV WASHINGTON.  
 PI Kaushansky K;  
 DR WPI; 95-292944/38.  
 DR P-PSDB; R82684.  
 PT Stimulation of erythropoiesis using thrombopoietin and opt.  
 PT erythropoietin - for the treatment of thrombocytopenia and anaemia.  
 PS Disclosure; Page 47-52; 66pp; English.  
 CC This sequence corresponds to a single allele of the human  
 CC thrombopoietin gene. Thrombopoietin stimulates erythropoiesis to  
 CC produce an increase in proliferation or differentiation of  
 CC erythroid cells or to increase reticulocyte counts at least 2-fold  
 CC over baseline reticulocyte counts and, optionally, platelet levels  
 CC to at least 20000/cu mm. The protein can be used in a composition,  
 CC optionally with erythropoietin, for use in the treatment of  
 CC thrombocytopenia and anaemia, such as that caused by destruction of  
 CC haematopoietic cells in bone marrow, in the treatment of cancer  
 CC with chemotherapy and radiation, and in pathological conditions  
 CC such as myelodysplasia, AIDS, aplastic anaemia, autoimmune disease  
 CC or inflammatory disease.  
 SQ Sequence 4823 BP; 1205 A; 1368 C; 1048 G; 1202 T;

Query Match 60.0%; Score 61.8; DB 1; Length 4823;  
 Best Local Similarity 80.0%; Pred. No. 5.2e-12;  
 Matches 72; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 7 tccaagctactcagaagactgaagcagaaggatcacttgaggccaggagttcaagatcag 66  
 ||| ||| | | || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 2917 TCCCAGCACTTTGGGAGGCTGAGGCAGGTGGATCACCTGAGGTCAGGAGTTCAAGATCAG 2858  
  
 Qy 67 cctgagcaacatagngaaaccctatctcta 96  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 2857 CCTGCCCAACATGGTGAAACCCCATCTCTA 2828



Query= SEQ ID NO:9  
(171 letters)

Sequences producing significant alignments:	Score (bits)	E Value
---	-----------------	------------

AC012640 ACCESSION:AC012640 NID: gi 27356677 gb AC012640.12 Ho...	<u>149</u>	2e-33
AC034241 ACCESSION:AC034241 NID: gi 17975241 gb AC034241.4 Hom...	<u>149</u>	2e-33

>AC012640 ACCESSION:AC012640 NID: gi 27356677 gb AC012640.12 Homo  
sapiens chromosome 5 clone CTD-2256P15, complete sequence  
Length = 145122

Score = 149 bits (75), Expect = 2e-33  
Identities = 124/137 (90%), Gaps = 2/137 (1%)  
Strand = Plus / Plus

Query: 20 tgtgaggacacagcnagaagcaagtctntgcatgncnagaagaacggcctcaacagacac 79  
|||||  
Sbjct: 54846 tgtgaggacacagcgagaagcaagtatctgcaagtcaagaagaaaggcctcaacagacac 54905

Query: 80 canncctgccagcaccttgatcttgg-cttntggcctccagaactgtgaaagantaaaga 138  
|| |||||  
Sbjct: 54906 cagccctgccagcaccttgatcttggacttctggcctccagaactgtgaaagaataaaa-a 54964

Query: 139 ttctgttggtttaagcca 155  
|||||  
Sbjct: 54965 ttctgttggtttaagcca 54981

>AC034241 ACCESSION:AC034241 NID: gi 17975241 gb AC034241.4 Homo  
sapiens chromosome 5 clone CTD-2360020, complete sequence  
Length = 77702

Score = 149 bits (75), Expect = 2e-33  
Identities = 124/137 (90%), Gaps = 2/137 (1%)  
Strand = Plus / Plus

Query: 20 tgtgaggacacagcnagaagcaagtctntgcatgncnagaagaacggcctcaacagacac 79  
|||||  
Sbjct: 10337 tgtgaggacacagcgagaagcaagtatctgcaagtcaagaagaaaggcctcaacagacac 10396

Query: 80 canncctgccagcaccttgatcttgg-cttntggcctccagaactgtgaaagantaaaga 138  
|| |||||  
Sbjct: 10397 cagccctgccagcaccttgatcttggacttctggcctccagaactgtgaaagaataaaa-a 10455

Query: 139 ttctgttggtttaagcca 155  
|||||  
Sbjct: 10456 ttctgttggtttaagcca 10472

SEQ ID No 10

T32454

ID T32454 standard; DNA; 30967 BP.

AC T32454;

DT 10-DEC-1996 (first entry)

DE Calpain large subunit 1 gene.

KW Calpain; subunit; calcium; protease; mutation; treatment;

KW detection; identification; diagnosis; limg girdle muscular dystrophy;

KW LGMD2; calcium activated neutral protease; CANP; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT exon 1367. .1674

FT /\*tag= a

FT /label= Exon 1.

FT intron 1675. .3689

FT /\*tag= b

FT /label= Intron 1.

FT exon 3690. .3759

FT /\*tag= c

FT /label= Exon 2.

FT intron 3760. .5390

FT /\*tag= d

FT /label= Intron 2.

FT exon 5391. .5509

FT /\*tag= e

FT /label= Exon 3.

FT intron 5510. .7015

FT /\*tag= f

FT /label= Intron 3.

FT exon 7016. .7050

FT /\*tag= g

FT /label= Exon 4.

FT intron 7051. .8128

FT /\*tag= h

FT /label= Intron 4.

FT exon 8129. .8297

FT /\*tag= i

FT /label= Exon 5.

FT intron 8298. .8889

FT /\*tag= j

FT /label= Intron 5.

FT exon 8890. .9297

FT /\*tag= k

FT /label= Exon 6.

FT intron 9298. .11843

FT /\*tag= l

FT /label= Intron 6.

FT exon 11844. .11927

FT /\*tag= m

FT /label= Exon 7.

FT intron 11928. .13458

FT /\*tag= n

FT /label= Intron 7.

FT exon 13459. .13545

FT /\*tag= o

FT /label= Exon 8.

FT intron 13456. .15026

FT /\*tag= p

FT /label= Intron 8.

FT	exon	15027. .16104
FT		/*tag= q
FT		/label= Exon 9.
FT	intron	16105. .17284
FT		/*tag= r
FT		/label= Intron 9.
FT	exon	17285. .17445
FT		/*tag= s
FT		/label= Exon 10.
FT	intron	17446. .19448
FT		/*tag= t
FT		/label= Intron 10.
FT	exon	19449. .19618
FT		/*tag= u
FT		/label= Exon 11.
FT	intron	19619. .19929
FT		/*tag= v
FT		/label= Intron 11.
FT	exon	19930. .19941
FT		/*tag= w
FT		/label= Exon 12.
FT	intron	19942. .20604
FT		/*tag= x
FT		/label= Intron 12.
FT	exon	20605. .20813
FT		/*tag= y
FT		/label= Exon 13.
FT	intron	20814. .21551
FT		/*tag= z
FT		/label= Intron 13.
FT	exon	21552. .21558
FT		/*tag= aa
FT		/label= Exon 14.
FT	intron	21589. .23746
FT		/*tag= ab
FT		/label= Intron 14.
FT	exon	23747. .23764
FT		/*tag= ac
FT		/label= Exon 15.
FT	intron	23765. .26027
FT		/*tag= ad
FT		/label= Intron 15.
FT	exon	26028. .26141
FT		/*tag= ae
FT		/label= Exon 16.
FT	intron	26142. .27119
FT		/*tag= af
FT		/label= Intron 16.
FT	exon	27120. .27197
FT		/*tag= ag
FT		/label= Exon 17.
FT	intron	27198. .27602
FT		/*tag= ah
FT		/label= Intron 17.
FT	exon	27603. .27660
FT		/*tag= ai
FT		/label= Exon 18.
FT	intron	27661. .27747
FT		/*tag= aj
FT		/label= Intron 18.

FT exon 27748. .27812  
 FT /\*tag= ak  
 FT /label= Exon 19.  
 FT intron 27813. .28244  
 FT /\*tag= al  
 FT /label= Intron 19.  
 FT exon 28245. .28313  
 FT /\*tag= am  
 FT /label= Exon 20.  
 FT intron 28314. .29404  
 FT /\*tag= an  
 FT /label= Intron 20.  
 FT exon 29405. .28483  
 FT /\*tag= ao  
 FT /label= Exon 21.  
 FT intron 28484. .28698  
 FT /\*tag= ap  
 FT /label= Intron 21.  
 FT exon 28699. .28815  
 FT /\*tag= aq  
 FT /label= Exon 22.  
 FT intron 28816. .29101  
 FT /\*tag= ar  
 FT /label= Intron 22.  
 FT exon 29102. .29260  
 FT /\*tag= as  
 FT /label= Exon 23.  
 FT intron 29261. .29562  
 FT /\*tag= at  
 FT /label= Intron 23.  
 FT exon 29563. .29589  
 FT /\*tag= au  
 FT /label= Exon 24.  
 PN WO9616175-A2.  
 PD 30-MAY-1996.  
 PF 21-NOV-1995; E04575.  
 PR 22-NOV-1994; EP-402668.  
 PA (ASFR-) ASSOC FR CONTRE MYOPATHIES.  
 PI Beckmann J, Richard I;  
 DR WPI; 96-268611/27.  
 DR P-PSDB; R99579.  
 PT Human novel Calpain large sub:unit 1 gene encoding a calcium  
 PT dependent protease - used to develop prods. for the diagnosis and  
 PT treatment of limb-girdle muscular dystrophy 2 disease  
 PS Claim 1; Figure 8; 66pp; English.  
 CC The calpain large subunit 1 gene located on chromosome 15 codes for  
 CC a calcium activated neutral protease (CANP3) belonging to the  
 CC calpain family. Mutations in the gene induce limb-girdle muscular  
 CC dystrophy (LGMD) 2 disease. The gene, and fragments of it, can be  
 CC used in the prevention, treatment, diagnosis and detection of a  
 CC predisposition to LGMD2 disease. A cDNA version of the gene is  
 CC described in T32455,  
 SQ Sequence 30967 BP; 7629 A; 7648 C; 7675 G; 8015 T;

Query Match 10.9%; Score 32; DB 1; Length 30967;  
 Best Local Similarity 62.5%; Pred. No. 1.2;  
 Matches 50; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 182 cttgcgcactgtgagtcacctggacgagttactccacctctctgaacctcctcctcacttg 241

```

      ||| | |||| || || |||| || || |||| || || ||
Db 3185 CTTACTAGCTGTGTGTCTTTGCACGAGTTTCTTAACCTCTCTGGGCCTCAGTTTCCTTAT 3244
Qy 242 cataatgggaaaaataatgg 261
      | || || || || |
Db 3245 CTGAAAAATAACAATGATAG 3264

```

RESULT 7

N90388

ID N90388 standard; cDNA; 5719 BP.

AC N90388;

DT 20-OCT-1989 (first entry)

DE cDNA encoding human platelet-derived growth factor receptor.

KW cDNA; human platelet derived growth factor receptor; agonist;

KW antagonist; drugs; wound healing; atherosclerosis;

KW cancer; genetic disorders; antibodies.

OS Homo sapiens (human)

FH Key Location/Qualifiers

FT cds 462

FT /\*tag= a

FT signal\_peptide 462..557

FT /\*tag= b

FT 500..524

PN EP-327369-A.

PD 09-AUG-1989.

PF 02-FEB-1989; 301021.

PR 02-FEB-1988; US-151414.

PA (REGC) Univ of California.

PI Williams L T, Escobedo J E;

DR WPI; 89-229378/32.

DR P-PSDB; P90646.

PT New DNA encoding human platelet derived growth factor receptor

PT - useful eg for assessing agonist and antagonist drugs.

PS Claim 1; page 3; 12pp; English.

CC cDNA encoding human platelet derived growth factor receptor (see P90646

CC for features). Used to make probes and antibodies, and to evaluate drugs.

SQ Sequence 5719 BP; 1266 A; 1714 C; 1548 G; 1191 T;

Query Match 10.0%; Score 29.4; DB 1; Length 5719;

Best Local Similarity 55.3%; Pred. No. 4.1;

Matches 57; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```

Qy 153 tgggtccttccaggacactgacgtctcagcttgcgactgtgagtcctggacgagttac 212
      || | | | | | | | | | | | | | | | | | | | | | |
Db 4633 TGTGCCAGTATATGGCCCTGGCTCTGCATTGGACCTGCTATGAGGCTTTGGAGGAATCCC 4692

Qy 213 tccacctctctgaacctcctcctcacttgcataatgggaaaaa 255
      || ||||| |||| | | | | | | | | | |
Db 4693 TCACCCTCTCTGGGCCTCAGTTTCCCCTTCAAAAAATGAATAA 4735

```

Query= SEQ ID NO:10  
(294 letters)

Sequences producing significant alignments:	Score (bits)	E Value
---	-----------------	------------

AC140062 ACCESSION:AC140062 NID: gi 29150317 gb AC140062.11 Ho...	<u>351</u>	5e-94
---	------------	-------

>AC140062 ACCESSION:AC140062 NID: gi 29150317 gb AC140062.11 Homo  
sapiens 12 BAC RP13-298C8 (Roswell Park Cancer Institute  
Human BAC Library) complete sequence  
Length = 64695

Score = 351 bits (177), Expect = 5e-94  
Identities = 180/181 (99%)  
Strand = Plus / Plus

Query: 114 aggcactgggtaggaacacagccaagaacgattgcaggatgggtccttccaggacactga 173  
|||||  
Sbjct: 1688 aggcactgggtaggaacacagccaagaacgattgcaggatgggtccttccaggacactga 1747

Query: 174 cgtctcagcttgcgccactgtgagtcacctggacgagttactccacctctctgaacctcctc 233  
|||||  
Sbjct: 1748 cgtctcagcttgcgccactgtgagtcacctggacgagttactccacctctctgaacctcctc 1807

Query: 234 ctcacttgcataatgggaaaaataatggacataggagatgaaacaagaccttgagacc 293  
|||||  
Sbjct: 1808 ctcacttgcataatgggaaaaataatggacataggagatgaaacaagaccttgagacc 1867

Query: 294 a 294  
|  
Sbjct: 1868 a 1868

Query= SEQ ID NO:11  
(241 letters)

Sequences producing significant alignments:	Score (bits)	E Value
AC112518.1.1.78409	<u>426</u>	e-117

>AC112518.1.1.78409  
Length = 78409

Score = 426 bits (215), Expect = e-117  
Identities = 232/239 (97%)  
Strand = Plus / Minus

Query: 3 atgccttctaaacagcctaccctgcccagngccatgattactgtgaccacatcttcagag 62  
|||||  
Sbjct: 2737 atgccttctaaacagcctaccctgccaagtgccatgattactgtgaccacatcttcagaa 2678

Query: 63 ccagaaaacaggatacctggccctaagcatgcactcatggagcanaagagttttaaactct 122  
|||||  
Sbjct: 2677 ccagaaaacaggatacctggccctaagcatgcactcatggagcagaagagttttaaactct 2618

Query: 123 gntatgccacagaagacagaagataaacatgcttactacacttgtnaagcaacatgcagcc 182  
|||  
Sbjct: 2617 ggaatgccacagaagacagaagataaacatgcttactacacttgtaaagcaacatgcagcc 2558

Query: 183 agccatttccagtgcaaattatctcattgcatagtggtgacaactaaagggtcataacccat 241  
|||||  
Sbjct: 2557 agccatttccagtgcaaattatctcattgcatagtggtgacaactaaagggtcataacccat 2499

Query= SEQ ID NO:12  
(197 letters)

Sequences producing significant alignments:	Score (bits)	E Value
---	-----------------	------------

AL158207 ACCESSION:AL158207 NID: gi 12717949 emb AL158207.15 H...	391	e-106
---	-----	-------

>AL158207 ACCESSION:AL158207 NID: gi 12717949 emb AL158207.15 Human DNA  
sequence from clone RP11-409K20 on chromosome 9 Contains  
the TOR1B gene for torsin family 1 member B (torsin B)  
(DQ1), the DYT1 gene for "dystonia 1, torsion" (autosomal  
dominant; torsin A) (DQ2, TOR1A), the gene for  
hepatocellular carcinoma-associated antigen 59 (HSPC220,  
LOC51759), the USP20 gene for ubiquitin specific protease  
20 (KIAA1003), and the gene for formin-binding protein 17  
(FBP17, includes KIAA0554, FLJ13619, FLJ10754 and  
FLJ10113). Contains ESTs, STSs, GSSs and four CpG islands,  
c

Length = 169963

Score = 391 bits (197), Expect = e-106  
Identities = 197/197 (100%)  
Strand = Plus / Plus

Query: 1	acaggatgcctgtaatcattattcagtgagcagcaacctgcagcagctcctcctgactgg	60
Sbjct: 137396	acaggatgcctgtaatcattattcagtgagcagcaacctgcagcagctcctcctgactgg	137455

Query: 61	cagatgggcctggcgccacccagaggctggggacacagcaagaatccagcacagcaccg	120
Sbjct: 137456	cagatgggcctggcgccacccagaggctggggacacagcaagaatccagcacagcaccg	137515

Query: 121	atcccgattccctcctcccccactacctgagccatggacctcattttgtggacaaaatt	180
Sbjct: 137516	atcccgattccctcctcccccactacctgagccatggacctcattttgtggacaaaatt	137575

Query: 181	aaacttgccactttcac	197
Sbjct: 137576	aaacttgccactttcac	137592



V02739

Query Match 8.5%; Score 32.8; DB 1; Length 3223;  
Best Local Similarity 61.9%; Pred. No. 0.33;  
Matches 52; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```

Qy      166  tctggtcaccaatttcacagcctggacagagcaagaaggtgcggtggttaggagggcg  225
          | | ||||| | | | | | | | | | | | | | | | | | | | | | | |
Db      2695  TGTCTGTCACCAAAGTCACCCGGCCGTGGAAGCGAGCAGCAGCAGCAAGCTGAGGAGGGCAA  2754

Qy      226  cctgccgggggggatcgtctgtcc  249
          | | ||| ||||| | | | |
Db      2755  GCAGATGGGAAGGATCGTCCGTGC  2778

```

RESULT 1  
V89805/c  
ID V89805 standard; cDNA; 265 BP.  
AC V89805;  
DT 15-FEB-1999 (first entry)  
DE EST clone CJ498.  
KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;  
KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;  
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;  
KW gene therapy; ss.  
OS Homo sapiens.  
PN WO9845436-A2.  
PD 15-OCT-1998.  
PF 10-APR-1998; U06955.  
PR 10-APR-1997; US-838821.  
PA (GEMY ) GENETICS INST INC.  
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,  
PI Racie LA, Spaulding V, Treacy M;  
DR WPI; 99-070077/06.  
PT New polynucleotides encoding human secreted proteins - derived from  
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries.  
PS Claim 1; Page 336; 618pp; English.  
CC The present sequence represents a human expressed sequence tag (EST).  
CC The polynucleotide, which is a secreted EST, and the encoded protein  
CC are predicted to have useful biological activities which would make  
CC them suitable for treating, preventing or ameliorating medical  
CC conditions in humans and animals, although no supporting data is  
CC given. Suggested activities include nutritional activity, immune  
CC stimulating or suppressing activity, haematopoiesis regulating  
CC activity, tissue growth activity, activin/inhibin activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
CC activity, receptor/ligand activity, anti-inflammatory activity,  
CC cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The polynucleotide may also be useful for gene therapy.  
SQ Sequence 265 BP; 62 A; 59 C; 59 G; 85 T;

Query Match 10.3%; Score 33.6; DB 1; Length 265;  
Best Local Similarity 61.4%; Pred. No. 0.1;  
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 14 ctcagcagacnaaccacagcttcctgccctttgcagatggcntgaanataagagtttgcc 73  
| | | | | | | | | | | | | | | | | | | | | |  
Db 110 CACTGCAGAAAAATTGAAGTTGAATCCCCCTCTGCACACAGACAGGGAATAAAAGTTTGAC 51  
Qy 74 aaacaactaagatgggctcttga 96  
| | | | | | | | | | | | | | | | | | | | | |  
Db 50 AAACACCCAAGAAAACTCTTTA 28

RESULT 1  
B17653/c  
LOCUS B17653 548 bp DNA GSS 04-JUN-1998  
DEFINITION 347F4.TPB CIT978SKA1 Homo sapiens genomic clone A-347F04,  
genomic survey sequence.  
ACCESSION B17653  
VERSION B17653.1 GI:2125402  
KEYWORDS GSS.

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 548)  
 AUTHORS Adams,M.D., Kelley,J.M., Rounsley,S.R. and Venter,J.C.  
 TITLE Use of a BAC End Sequence Database for Sequence-Ready Map Building  
 JOURNAL Unpublished (1997)  
 COMMENT On Dec 15, 1999 this sequence version replaced gi:4575579.  
 Other\_GSSs: 347F04.TVB  
 Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadams@tigr.org  
 Clones are available from Research Genetics (info@resgen.com). BAC  
 end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
 Seq primer: SP6  
 Class: BAC ends.  
 FEATURES Location/Qualifiers  
 source 1. .548  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="A-347F04"  
 /clone\_lib="CIT978SKA1"  
 /sex="Female"  
 /cell\_type="Fibroblast"  
 /note="Vector: pBAC108L; Site\_1: HindIII; Site\_2: HindIII;  
 CalTech Human BAC Library A1"  
 BASE COUNT 153 a 109 c 102 g 184 t  
 ORIGIN

Query Match 85.8%; Score 279.6; DB 120; Length 548;  
 Best Local Similarity 90.5%; Pred. No. 2.2e-73;  
 Matches 296; Conservative 0; Mismatches 30; Indels 1; Gaps 1;

```

Qy      1 ggacagtggctaactcagcagacnaaccacagcttctctgccctttgcagatggcntgaan 60
      |||
Db      512 GGACAGTGGCTAACTCAGCAGACGAACCAGAGCTTCATGCCCTTTGCAGATGGCATGAAG 453

Qy      61 ataagagtttgccaaacaactaagatgggctcttgattgagcaaanaaaccacaacatgg 120
      |||
Db      452 ATAAGAGTTTGCCAAACAACCTAAGATGGGCTCTTGATTGAGCAAAGAAACCACAACATGG 393

Qy      121 gacacacagagccaccctattgncctactgtcattcaagcttaaaggagacatatctaca 180
      |||
Db      392 GACACACAGAGCCACCTAATTGCCATACTGTCAATCAAGCTTAAAGGAGACATATCTACA 333

Qy      181 gacagggtttgagcctagtnatggnganaactttcttgatgtctcaacancctgganat 240
      |||
Db      332 GACAGGGTTTGAGCATAGTAATGGTGAGAACTTTCTTGATGTCTCAACAGCCTGGAGAT 273

Qy      241 gannntcccnacaaggcagaanananchaggtggnacattgntnntattgctttttatt-ca 299
      ||
Db      272 GAAATTTCCAAGAAGGCAGAAAATAGAGGTGGCACATTGGTTTTATTGTTTTTTATTACA 213

```

Qy 300 attataaaaagtaatgcatgcttttttgt 326  
|||||  
Db 212 ATTATAAAAAGTAATGCATGCTTTTGT 186

Query= SEQ ID NO:13  
(387 letters)

Sequences producing significant alignments:	Score (bits)	E Value
---	-----------------	------------

AL161936.15.1.155584	<u>753</u>	0.0
----------------------	------------	-----

>AL161936.15.1.155584  
Length = 155584

Score = 753 bits (380), Expect = 0.0  
Identities = 385/387 (99%)  
Strand = Plus / Plus

Query: 1	tggtgcttactaaaaattgaataancgtggaaaagagaaaatctccctctttaaaaggaa	60
Sbjct: 144402	tggtgcttactaaaaattgaataaacgtggaaaagagaaaatctccctctttaaaaggaa	144461

Query: 61	cactgttgtggacattttaaaatgcaaacgccttggctggaagtcagaaatcgtgttctc	120
Sbjct: 144462	cactgttgtggacattttaaaatgcaaacgccttggctggaagtcagaaatcgtgttctc	144521

Query: 121	tctgctaaacctggtgtagcatttaacacgcttgaagtggagggcatctggtcaccaattt	180
Sbjct: 144522	tctgctaaacctggtgtagcatttaacacgcttgaagtggagggcatctggtcaccaattt	144581

Query: 181	cacagcctggacagagcaagaaggtgCGGCTGGCTtaggaggCGGCTGCCGGGGGGGAT	240
Sbjct: 144582	cacagcctggacagagcaagaaggtgCGGCTGGCTtaggaggCGGCTGCCGGGGGGGAT	144641

Query: 241	cgtctgtccatctgggcttggtaaatgtcaagggtcatttccctgtcctgacatttgatt	300
Sbjct: 144642	cgtctgtccatctgggcttggtaaatgtcaagggtcatttccctgtcctgacatttgatt	144701

Query: 301	gtgaagcagggttgcgaggtaactctttcaagggactggactgtgacagtcaccatagttg	360
Sbjct: 144702	gtgaagcagggttgcgaggtaactctttcaagggactggactgtgacagtcaccatagttg	144761

Query: 361	gacaataaaaacccgaacatccttcacc	387
Sbjct: 144762	gacaataaaaacccgaacatccttcacc	144788

Query= SEQ ID NO:14  
(326 letters)

Sequences producing significant alignments:	Score (bits)	E Value
AC092768 ACCESSION:AC092768 NID: gi 18182777 gb AC092768.6 Hom...	<u>466</u>	e-128

>AC092768 ACCESSION:AC092768 NID: gi 18182777 gb AC092768.6 Homo  
sapiens chromosome 11, clone RP11-1149L18, complete  
sequence  
Length = 146364

Score = 466 bits (235), Expect = e-128  
Identities = 301/327 (92%), Gaps = 1/327 (0%)  
Strand = Plus / Minus

Query: 1 ggacagtggctaactcagcagacnaaccacagcttcctgccctttgcagatggcntgaan 60  
|||||  
Sbjct: 8644 ggacagtggctaactcagcagacgaaccagagcttcctgccctttgcagatggcatgaag 8585

Query: 61 ataagagtttgccaaacaactaagatgggctcttgattgagcaanaaaccacaacatgg 120  
|||||  
Sbjct: 8584 ataagagtttgccaaacaactaagatgggctcttgattgagcaaaagaaaccacaacatgg 8525

Query: 121 gacacacagagccaccctattgncctactgtcattcaagcttaaaggagacatatctaca 180  
|||||  
Sbjct: 8524 gacacacagagccaccctattgccctactgtcattcaagcttaaaggagacatatctaca 8465

Query: 181 gacagggtttgagcctagtnatggnganaactttcttgatgtctcaacancctgganat 240  
|||||  
Sbjct: 8464 gacagggtttgagcctagtaatggtgagaactttcttgatgtctcaacagcctggagat 8405

Query: 241 ganmntcccnacaaggcagaaanancnaggtggnacattgntnntattgctttttatt-ca 299  
|| |||||  
Sbjct: 8404 gaaattcccaagaaggcagaaaatagaggtggcacattgggttttattgttttttattaca 8345

Query: 300 attataaaaagtaatgcatgctttttgt 326  
|||||  
Sbjct: 8344 attataaaaagtaatgcatgctttttgt 8318

# SEQ ID No 15

Q91200/c

ID Q91200 standard; cDNA; 3320 BP.  
AC Q91200;  
DT 11-DEC-1995 (first entry)  
DE H-NUC retinoblastoma protein binding protein.  
KW H-NUC; tumour suppressor; retinoblastoma binding protein;  
KW therapeutic; gene therapy; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT cds 101. .2576  
FT /\*tag= a  
PN WO9517198-A1.  
PD 29-JUN-1995.  
PF 20-DEC-1994; U14813.  
PR 20-DEC-1993; US-170586.  
PA (TEXA ) UNIV TEXAS SYSTEM.  
PI Chen P, Lee W;  
DR WPI; 95-240467/31.  
DR P-PSDB; R75848.  
PT DNA encoding a retinoblastoma protein binding protein - used in the  
PT gene therapy of cancers, esp. breast cancer.  
PS Claim 4; Fig 3A-3I; 85pp; English.  
CC The H-NUC DNA and protein encoded by it may be used to suppress the  
CC neoplastic phenotype of a cancer cell which lacks endogenous H-NUC  
CC protein. The DNA and protein inhibit cancer, especially mamma  
CC carcinoma, cell division and proliferation. A retro virus vector or  
CC adeno virus vector (AC-H-NUC) may be used for the ex vivo gene  
CC therapy of cancer, where the H-NUC gene is transferred to abnormally  
CC proliferating cells; gene expression in sufficient amounts  
CC suppresses proliferation of those cells. The cells are then  
CC returned to the original mammal.  
SQ Sequence 3320 BP; 1049 A; 674 C; 662 G; 935 T;

Query Match 16.4%; Score 27.2; DB 1; Length 3320;  
Best Local Similarity 72.9%; Pred. No. 6.4;  
Matches 35; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 106 tcttcaagtgcttgtaaggccatttgtctatttctactctcaagtaaa 153  
||||||| ||||| ||| | || | | ||||| | ||| |||||  
Db 2283 TCTTCAAGTTCTTGTAAGCAGACTTATATTTTTCATTTGCAAATAAA 2236

## RESULT 7

Q59800

ID Q59800 standard; cDNA; 372 BP.  
AC Q59800;  
DT 16-MAR-1994 (first entry)  
DE Human brain Expressed Sequence Tag EST00733.  
KW Gene transcription product; genetic markers; tagging; in vivo;  
KW transcription; mapping; locations; chromosomes; chromosomal; ss.  
OS Homo sapiens.  
PN WO9316178-A.  
PD 19-AUG-1993.  
PF 12-FEB-1993; U01294.  
PR 12-FEB-1992; US-837195.  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
PI Adams MD, Moreno RF, Venter CJ;

DR WPI; 93-272882/34.  
PT Enriched oligonucleotides and corresp. sequences - used as  
PT markers for human genes transcribed in-vivo, facilitate tagging  
PT of most human genes  
PS Example 4; Page 233; 500pp; English.  
CC The Expressed Sequence Tag was isolated from a human brain cDNA  
CC library as part of a large set of ESTs which can be used as markers  
CC for human genes transcribed in vivo. They can be used to facilitate  
CC tagging of most human genes, for mapping locations of expressed genes  
CC on chromosomes, for individual or forensic identification, for mapping  
CC locations of disease-associated genes, for identification of tissue  
CC type, and for prepn. of antisense sequences, probes and constructs.  
CC EST00733 has a "poor" coding probability as evaluated using the  
CC coding-region prediction program CRM. See also Q59041-Q61440.  
SQ Sequence 372 BP; 87 A; 93 C; 55 G; 134 T;

Query Match 16.3%; Score 27; DB 1; Length 372;  
Best Local Similarity 66.1%; Pred. No. 2.9;  
Matches 39; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 108 ttcaagtgccttgtaaggccatttgtctatttcactctcaagtaaataaaaaatattttt 166  
||| | ||||| | ||| | |||| | ||| | | |||| | |||||  
Db 151 TTCTACTGCTTGTTCAATACATCTCTCTATGTAAATCTTGACTCCATAATGAGGTTTTT 209



Query= SEQ ID NO:15  
(166 letters)

Sequences producing significant alignments:	Score (bits)	E Value
---	-----------------	------------

321 7e-86

```
>AC008115.3.1.158431
```

Length = 158431

Score = 321 bits (162), Expect = 7e-86

Identities = 165/166 (99%)

Strand = Plus / Minus

Query: 1      tcagtatcctgacctggcaaggtgttccttaacctccccctctggatcccccttagcacac 60

\_\_\_\_\_

Sbjct: 43020 tcagtatcctgacctggcaaggtgttccttaacctcccctctggatcccccttagcacac 42961

Query: 61 atctgggacaatggagcgttcagcaccacggacagcattacaccctcttcaagtgcttgt 120

\_\_\_\_\_

Sbjct: 42960 atctgggacaatggagcgttcagcaccacggacagcattacaccctcttcaagtgcttgt 42901

Query: 121 taaggccatttgtctatttcactctcaagtaaataaaaaatattttt 166

\_\_\_\_\_

Sbjct: 42900 taaagccatttgtctatttcactctcaagtaaataaaaaatatattttt 42855

SEQ ID NO 16

HSAC002043/c

LOCUS HSAC002043 226841 bp DNA HTG 30-APR-1997

DEFINITION Homo sapiens clone 381E11, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 4 unordered pieces.

ACCESSION AC002043

VERSION AC002043.1 GI:2062147

KEYWORDS HTG; HTGS\_PHASE1.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 226841)

AUTHORS Adams,M.D., Loftus,B.J., Zhou,L., Phillips,C., Brandon,R.C., Fuhrmann,J., Kim,U.J., Kerlavage,A.R. and Venter,J.C.

TITLE Human chromosome 16p13 BAC clone CIT987SK-381E11

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 226841)

AUTHORS Adams,M.D. and Loftus,B.J.

TITLE Direct Submission

JOURNAL Submitted (29-APR-1997) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA

REFERENCE 3 (bases 1 to 226841)

AUTHORS Adams,M.D.

TITLE Direct Submission

JOURNAL Submitted (30-APR-1997) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA

COMMENT

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 2951: contig of 2951 bp in length  
\* gap of unknown length  
\* 2952 14471: contig of 11520 bp in length  
\* gap of unknown length  
\* 14472 135316: contig of 120845 bp in length  
\* gap of unknown length  
\* 135317 226841: contig of 91525 bp in length.

FEATURES

Location/Qualifiers

source

1. .226841  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="381E11"

BASE COUNT 61420 a 49870 c 51181 g 64328 t 42 others

ORIGIN

Query Match 9.0%; Score 57.2; DB 54; Length 226841;  
Best Local Similarity 87.3%; Pred. No. 1e-07;  
Matches 62; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 568 agtagagacaagttttcgccatgttgggtcaagctgggtctcaaacttctaacctnacgtaa 627

|||||

Db 135306 AGTAGAGACAAAGGTTTCGCCATGTTGGTGAAGCTGGTCTCAAACCCCTGACCTCAGGTAA 135247



SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 274)

AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,  
Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,  
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,  
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,  
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,  
Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.  
and Marra,M.

TITLE Generation and analysis of 280,000 human expressed sequence tags

JOURNAL Genome Res. 6 (9), 807-828 (1996)

MEDLINE 97044478

COMMENT On May 7, 1998 this sequence version replaced gi:3119472.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 1750  
High quality sequence stops: 214  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1750 Std Error: 0.00  
Seq primer: M13RP1  
High quality sequence stop: 214.

FEATURES Location/Qualifiers

source 1. .274  
/organism="Homo sapiens"  
/db\_xref="GDB:3785124"  
/db\_xref="taxon:9606"  
/clone="IMAGE:212283"  
/clone\_lib="Soares fetal liver spleen 1NFLS"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)  
with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
1st strand cDNA was primed with a Pac I - oligo(dT) primer  
[5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 53 a 82 c 66 g 73 t

ORIGIN

Query Match 8.9%; Score 56.6; DB 86; Length 274;  
Best Local Similarity 86.1%; Pred. No. 1.8e-05;  
Matches 62; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 567 cagtagagacaagttttcgccatgttggtcaagctggtctcaaacttctaacctnacgta 626  
||||||| | ||||||| ||||||| ||||||| || |||| | |  
Db 167 CAGTAGAGACAGGGTTTCGCCATGTTGGTCAGGCTGGTCTCAAACCTCCTGACCTCAGGTG 226

Qy 627 atccaccccgct 638  
||||||| ||  
Db 227 ATCCACCCGCCT 238

RESULT 14

HSAC002043/c

LOCUS HSAC002043 226841 bp DNA HTG 30-APR-1997

DEFINITION Homo sapiens clone 381E11, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 4 unordered pieces.

ACCESSION AC002043

VERSION AC002043.1 GI:2062147

KEYWORDS HTG; HTGS\_PHASE1.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 226841)

AUTHORS Adams,M.D., Loftus,B.J., Zhou,L., Phillips,C., Brandon,R.C., Fuhrmann,J., Kim,U.J., Kerlavage,A.R. and Venter,J.C.

TITLE Human chromosome 16p13 BAC clone CIT987SK-381E11

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 226841)

AUTHORS Adams,M.D. and Loftus,B.J.

TITLE Direct Submission

JOURNAL Submitted (29-APR-1997) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA

REFERENCE 3 (bases 1 to 226841)

AUTHORS Adams,M.D.

TITLE Direct Submission

JOURNAL Submitted (30-APR-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

COMMENT \* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 2951: contig of 2951 bp in length  
\* gap of unknown length  
\* 2952 14471: contig of 11520 bp in length  
\* gap of unknown length  
\* 14472 135316: contig of 120845 bp in length  
\* gap of unknown length  
\* 135317 226841: contig of 91525 bp in length.

FEATURES Location/Qualifiers

source 1..226841  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="381E11"

BASE COUNT 61420 a 49870 c 51181 g 64328 t 42 others

ORIGIN

Query Match 9.0%; Score 57.2; DB 54; Length 226841;  
Best Local Similarity 87.3%; Pred. No. 1e-07;  
Matches 62; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 568 agtagagacaagtttttcgccatggttggtcaagctggtctcaaacttctaacctnacgtaa 627  
|||||  
Db 135306 AGTAGAGACAAGGTTTCGCCATGTTGGTGAAGCTGGTCTCAAACCCCTGACCTCAGGTAA 135247  
Qy 628 tccaccccgct 638  
|||||  
Db 135246 TCCACCCGCCT 135236







Qy 64 cagcctgagcaacatagnaaaccctatctctaaaaata 102  
||||| ||||||| ||||||| ||||| |||||||  
Db 65 CAGCCTAAGCAACATAGTGAAACCCCATCTCCAAAAATA 27

Sequences producing significant alignments:	Score (bits)	E Value
AL021391 ACCESSION:AL021391 NID: qi 4467344 emb AL021391.2 HS10...	347	2e-92

Score = 347 bits (175), Expect = 2e-92  
Identities = 175/175 (100%)  
Strand = Plus / Minus

Score = 248 bits (125), Expect = 1e-62  
Identities = 127/128 (99%)  
Strand = Plus / Minus

```
Query: 330      agtatgagaaaagaacaggctagattatgttatccagaacttcacaaccatcagatctag 389
              |||||
Sbjct: 90571    aqtatgagaaaagaacaggctagattatgttatccagaacttcacaaccatcagatctag 90512
```

Query: 390 acagaagg 397  
|||||||  
Sbjct: 90511 acagaagg 90504

Score = 111 bits (56), Expect = 2e-21  
Identities = 64/67 (95%)  
Strand = Plus / Minus

Query: 568 agtagagacaagttttcgccatggttggtcaagctggtctcaaacttctaacctnacgtaa 627  
|||||||  
Sbjct: 89080 agtagagacaagttttcgccatggttggtcaggctggtctcaaactcctaacctcacgtaa 89021

Query: 628 tccaccc 634  
|||||||  
Sbjct: 89020 tccaccc 89014

Score = 75.8 bits (38), Expect = 1e-10  
Identities = 46/50 (92%)  
Strand = Plus / Minus

Query: 219 ccagggttnnagtgattcccgtgcttcngnctcctgagaagctgggattac 268  
|||||||  
Sbjct: 94134 ccagggttcaagtgattcccgtgcttcagcctcctgagaagctgggattac 94085

Query= SEQ ID NO:17  
(403 letters)

Sequences producing significant alignments:

	Score (bits)	E Value
--	-----------------	------------

AC015933.9.1.249021	<u>668</u>	0.0
---------------------	------------	-----

>AC015933.9.1.249021  
Length = 249021

Score = 668 bits (337), Expect = 0.0  
Identities = 383/402 (95%), Gaps = 3/402 (0%)  
Strand = Plus / Plus

Query: 3	aaagagaaaaacaacattcaacancaacancaattttcccgaggatccctgcccacattca	62
Sbjct: 224797	aaagagaaaaacaacaa-caacaacaacaacaattttcccgaggatccctgcccacattca	224855

Query: 63	nagt-gncacatttacctacttnanaggggagatnaaagccnccactctaaggctccttat	121
Sbjct: 224856	gagtag-cagatttacctacttcaaagtggagatcaaagccacactctaaggctccttat	224914

Query: 122	ttccacaggctggnaagcaaacanggcntacaggctttgcangagtgtatcctaattctc	181
Sbjct: 224915	ttccacaggctggcaagcaaacaggcatacaggctttgcaagagtgtatcctaattctc	224974

Query: 182	ttactgaagaaaagtcaacagcagagacancacagaaaaaggaatcaaagaggccaaatc	241
Sbjct: 224975	ttactgaagaaaagtcaacagcagagacaacacagaaaaaggaatcaaagaggccaaatc	225034

Query: 242	tgnggactcaaaacaataagaaaaaataaatcaactttgctaaaaatttaagaatgccagg	301
Sbjct: 225035	tgtggactcaaaacaataagaaaaaataaatcaactttgctaaaaatttaagaatgccagg	225094

Query: 302	ggggtaggtaaatgcactgggaagtatgtgtggactatgatgataataaatctcctttca	361
Sbjct: 225095	ggggtaggtaaatgcactgggaagtatgtgtggactatgatgataataaatctcctttca	225154

Query: 362	atacaactgatatttatcagaccttgaataaaacactgaatg	403
Sbjct: 225155	atacaactgatatttatcagaccttgaataaaacactgaatg	225196

Query= SEQ ID NO:18  
(103 letters)

Sequences producing significant alignments:	Score (bits)	E Value
AL360270 ACCESSION:AL360270 NID: gi 11121069 emb AL360270.18 H...	<u>198</u>	1e-48

>AL360270 ACCESSION:AL360270 NID: gi 11121069 emb AL360270.18 Human DNA  
sequence from clone RP11-96K19 on chromosome 1, complete  
sequence  
Length = 172805

Score = 198 bits (100), Expect = 1e-48  
Identities = 102/103 (99%)  
Strand = Plus / Plus

Query: 1 actttctccaagctactcagaagactgaagcagaaggatcacttgaggccaggagttcaa 60  
|||||  
Sbjct: 93618 actttctccaagctactcagaagactgaagcagaaggatcacttgaggccaggagttcaa 93677

Query: 61 gatcagcctgagcaacatagngaaaccctatctctaaaaatac 103  
|||||  
Sbjct: 93678 gatcagcctgagcaacatagtgaaaccctatctctaaaaatac 93720